

Rec'd

PCT/JP2004/008051

日 本 国 特 許 庁
JAPAN PATENT OFFICE

03.06.2004

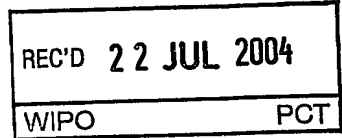
別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2003年 6月11日

出 願 番 号
Application Number: 特願2003-165946
[ST. 10/C]: [JP2003-165946]

出 願 人
Applicant(s): 松下電器産業株式会社

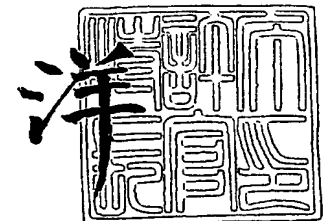


PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b).

2004年 7月 9日

特許庁長官
Commissioner,
Japan Patent Office

小 川



出証番号 出証特2004-3059493

【書類名】 特許願

【整理番号】 2033850102

【あて先】 特許庁長官殿

【国際特許分類】 G10L

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 鈴木 哲

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 金森 丈郎

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 河村 岳

【特許出願人】

【識別番号】 000005821

【氏名又は名称】 松下電器産業株式会社

【代理人】

【識別番号】 100109210

【弁理士】

【氏名又は名称】 新居 広守

【手数料の表示】

【予納台帳番号】 049515

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0213583

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声区間検出方法および装置

【特許請求の範囲】

【請求項1】 入力音響信号から音声が含まれる区間を音声区間として検出する音声区間検出方法であって、

前記入力音響信号に対し、あらかじめ定められた時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出ステップと、

前記音響特徴量のフレーム間における持続性を評価し、評価結果に従って音声区間を決定する区間決定ステップとを含む

ことを特徴とする音声区間検出方法。

【請求項2】 前記音響特徴量抽出ステップでは、前記入力音響信号に対しフレーム単位で周波数変換を行ない、前記音響特徴量より調波構造を抽出する

ことを特徴とする請求項1に記載の音声区間検出方法。

【請求項3】 前記音響特徴量抽出ステップでは、さらに、前記周波数変換の結果より調波構造を強調し、前記音響特徴量を抽出する

ことを特徴とする請求項2に記載の音声区間検出方法。

【請求項4】 前記音響特徴量抽出ステップでは、さらに、前記周波数変換の結果より調波構造を抽出し、当該調波構造を含む所定の帯域の周波数変換の結果を、前記音響特徴量とする

ことを特徴とする請求項3に記載の音声区間検出方法。

【請求項5】 前記区間決定ステップでは、前記音響特徴量の同一あるいは異なるフレームの音響特徴量間における相関値に基づいて、前記持続性を評価する

ことを特徴とする請求項1～4のいずれかに記載の音声区間検出方法。

【請求項6】 前記区間決定ステップは、
前記音響特徴量のフレーム間における持続性を評価する評価値を算出する評価ステップと、

前記評価値の時間的な連続性を評価し、評価結果に従って音声区間を決定する音声区間決定ステップとを含む

ことを特徴とする請求項 1～4 のいずれかに記載の音声区間検出方法。

【請求項 7】 前記区間決定ステップは、さらに、

所定数のフレームにわたる、音響特徴量抽出ステップにおいて算出される音響特徴量または、前記評価ステップにおいて算出される前記評価値と、第 1 の所定しきい値との比較に基づいて、前記入力音響信号の音声雑音比を推定するステップと、

推定された前記音声雑音比が第 2 の所定しきい値以上の場合には、前記評価ステップにおいて算出される前記評価値に基づいて前記音声区間を決定するステップとを含み、

前記音声区間決定ステップでは、前記音声雑音比が前記第 2 の所定しきい値未満の場合に、前記評価値の時間的な連続性を評価し、評価結果に従って前記音声区間を決定する

ことを特徴とする請求項 6 に記載の音声区間検出方法。

【請求項 8】 前記区間決定ステップでは、2 種以上の異なる間隔のフレーム間における複数の相関値に基づいて、前記持続性を評価する

ことを特徴とする請求項 1～4 のいずれかに記載の音声区間検出方法。

【請求項 9】 前記区間決定ステップでは、前記入力音響信号の音声雑音比に基づいて、前記 2 種以上の異なる間隔のフレーム間における相関値のいずれかを選択し、選択された相関値に基づいて前記持続性を評価する

ことを特徴とする請求項 8 に記載の音声区間検出方法。

【請求項 10】 前記区間決定ステップでは、前記音響特徴量のフレーム間における相関値と、前記相関値を所定フレーム数にわたり平均した平均値との補正相関値に基づいて、前記持続性を評価する

ことを特徴とする請求項 1～4 のいずれかに記載の音声区間検出方法。

【請求項 11】 入力音響信号から音声が含まれる区間を音声区間として検出する音声区間検出装置であって、

前記入力音響信号に対し、あらかじめ定められた時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出手段と、

前記音響特徴量のフレーム間における持続性を評価し、評価結果に従って音声

区間を決定する区間決定手段とを含む

ことを特徴とする音声区間検出装置。

【請求項 12】 入力音響信号に含まれる音声を認識する音声認識装置であって、

前記入力音響信号に対し、あらかじめ定められた時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出手段と、

前記音響特徴量のフレーム間における持続性を評価し、評価結果に従って音声区間を決定する区間決定手段と、

前記区間決定手段で決定された音声区間において音声認識を行なう認識手段とを備える

ことを特徴とする音声認識装置。

【請求項 13】 入力音響信号に含まれる音声を録音する音声録音装置であって、

前記入力音響信号に対し、あらかじめ定められた時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出手段と、

前記音響特徴量のフレーム間における持続性を評価し、評価結果に従って音声区間を決定する区間決定手段と、

前記区間決定手段で決定された音声区間における入力音響信号を録音する録音手段とを備える

ことを特徴とする音声録音装置。

【請求項 14】 コンピュータに請求項 1～10 のいずれかに記載の音声区間検出方法のステップを実行させるためのプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、入力音響信号から音声が含まれる区間を音声区間として検出する音声区間検出方法に関し、特に、環境雑音下における音声区間検出方法に関する。

【0002】

【従来の技術】

人間の音声は、声帯の振動と発声器官の共振によって形成されており、音の大きさと音の高低を区別するために声帯を制御して振動の周波数を変化させたり、鼻や舌などの発声器官の位置つまり声道形状を変動させることで、人はさまざまな音を発声していることが知られている。このように生成される音声を、音響信号として捕えると、その特徴は、周波数とともに緩やかに変化する成分である、スペクトル包絡と、短時間の周期的（有声音母音などの場合）にまたは非周期的に変化する成分（子音や無声音母音の場合）である、スペクトル微細構造から構成されていることが知られている。前者のスペクトル包絡成分が発声器官の共振特性を表しており、人間の喉や口の形をあらわす特徴量として用いられ、たとえば音声認識の特徴量としても用いられている。一方、後者のスペクトル微細構造は、音源の周期性を表しており、声帯の基本周期（ピッチ）、音の高低を表す特徴量として用いられている。音声信号のスペクトルは、これら2つの要素の積で表現されている。とくに母音部などにおいて、後者の基本周期およびその高調波成分をよく残している信号は、音声の調波構造とも呼ばれている。

【0003】

従来、入力音響信号から音声区間を検出する手法は、様々提案されている。それらを大きく分類すると、入力音響信号の帯域パワーやスペクトルの概形を示すスペクトル包絡などの振幅情報を用いて識別する方法（以下、「方法1」という。）、口映像を動画像解析することにより、その開閉を検出する方法（以下、「方法2」という。）、音声や雑音を表現する音響モデルと入力音響信号の音響特徴量とを比較することにより音声区間を検出する方法（以下、「方法3」という。）、および音声の調音器官の特徴である声道形状によって形成されるスペクトル包絡形状や声帯振動によって形成される調波構造に着目して音声区間を決定する方法（以下、「方法4」という。）などがある。

【0004】

しかし、方法1では、もともと振幅情報だけで音声と雑音とを識別することが難しいという問題を含んでいる。このため、方法1では、音声区間と雑音区間とを仮定し、音声区間と雑音区間とを区別するために設定したしきい値を再学習することにより、音声区間の検出を行なっている。したがって、学習過程において

雑音区間の振幅が音声区間の振幅に対して大きくなる（すなわち音声雑音比（以下、「SNR」という。）が0 dB程度まで低下する）と、雑音区間であるか音声区間であるかの仮定そのものの精度が性能に影響し、しきい値学習の精度が劣化してしまう。その結果として、音声区間検出の性能が劣化するという問題がある。

【0005】

また、方法2では、例えば音入力を用いずに画像だけを用いて口が開いたことを検出するようにすれば、その音声区間検出推定精度は、SNRとは無関係に一定に保つことが可能である。しかし、画像解析処理は音声信号の解析処理に比べて、コストが高いことと、口がカメラの方向に向いていない場合には音声区間の検出ができないという問題がある。

【0006】

さらに、方法3では、想定した環境雑音下での性能は確保されるものの、雑音を想定することそのものが難しいため、この方法を使用できる環境は限定的となってしまう。その場の雑音環境を学習する手法も提案されているが、振幅情報を利用する方法（方法1）と同様に、学習方法の精度に依存して性能が劣化するという問題もある。

【0007】

一方、音声の調音器官の特徴である、声道形状によって形成されるスペクトル包絡形状や声帯振動によって形成される調波構造に着目して音声区間を決定する方法（方法4）も提案されてきた。

【0008】

スペクトル包絡形状を利用した方法には、帯域パワー例えばケプストラムの連続性を評価する方法などがあるが、SNRが低下した状況では雑音のオフセット成分との区別がつきにくくなるため、性能が劣化する。

【0009】

調波構造に着目した方法として、ピッチ検出法はその手法の一つであり、時間軸上の自己相関や高次ケフレンシーを抽出する方法、周波数軸上の自己相関を行う方法等が提案されている。しかし、これらの方法は、対象とする信号が単一の

ピッチ（高調波の基本周波数）を持つ信号でない場合には音声区間の抽出が困難であり、環境雑音によって抽出誤りが発生し易い等の問題がある。

【0010】

また、複数種類の音響信号が混在した音響信号から、人の音声や特定の楽器音等の調波構造を持った音響信号を強調したり、抑圧したり、分離抽出したりする技術が知られている。例えば音声信号に対しては、雑音と音声信号とが混在した音響信号から雑音のみを抑圧する雑音抑圧装置（たとえば、特許文献1参照。）が、また音楽に対しては演奏に含まれる旋律の分離方法や除去方法（たとえば、特許文献2参照。）が、それぞれ提案されている。

【0011】

しかし、特許文献1に記載の方法では、入力信号の線形予測残差信号を帯域ごとに観察することで音声および非音声の検出を行っている。したがって、線形予測がうまく機能しない低SNRの非定常雑音下では性能が劣化するという問題がある。

【0012】

また、特許文献2に記載の方法は、同一の音程の音が一定時間持続するという音楽の旋律特有の性質を利用した方法である。このため、この方法を、音声と雑音との区別にそのまま用いることは困難であるという問題がある。

【0013】

調波構造を表現する音響特徴量そのものを評価関数に用いる手法（たとえば、特許文献3参照。）も提案されている。図9は、特許文献3で提案されている方法を用いた音声区間決定装置の概略構成を示すブロック図である。

【0014】

図9に示される音声区間検出装置10は、入力信号中の音声区間を決定する装置であり、FFT（Fast Fourier Transform）部100と、調波構造評価部101と、調波性ピーク検出部102と、ピッチ候補検出部103と、フレーム間振幅差分調波性評価部104と、音声区間決定部105とを備える。

【0015】

FFT部100は、入力信号に対し、フレーム（たとえば、1フレームは、1

0 msec) ごとにFFT処理を行ない、入力信号を周波数変換し、各種の分析を行なう。調波構造評価部101は、FFT部100より得られた周波数分析結果より、フレームごとに調波構造性を有するか否かの評価を行なう。調波性ピーク検出部102は、調波構造評価部101で抽出された調波構造をローカルピーク形状に変換し、ローカルピークを検出する。

【0016】

ピッチ候補検出部103は、調波性ピーク検出部102で検出されたローカルピークを時間軸方向（フレーム方向）にトラッキングすることによりピッチ検出を行なう。ピッチとは、調波構造の基本周波数のことである。

【0017】

フレーム間振幅差分調波性評価部104は、FFT部100における周波数分析の結果得られた振幅をフレーム間で差分し、差分値を求め、その差分値より着目しているフレームが調波構造を有するか否かの評価を行なう。

【0018】

音声区間決定部105は、ピッチ候補検出部103で検出されたピッチと、フレーム間振幅差分調波性評価部104の評価結果とを総合的に判断し、音声区間を決定する。

したがって、図9に示される音声区間検出装置10では、単一のピッチのみを有する音響信号のみならず、複数のピッチを有する音響信号であっても、音声区間を決定できる。

【0019】

【特許文献1】

特開平9-153769号公報

【0020】

【特許文献2】

特開平11-143460号公報

【0021】

【特許文献3】

特開2001-222289号公報

【0022】

【発明が解決しようとする課題】

しかしながら、ピッチ候補検出部103において、ローカルピークをトラッキングする際には、ローカルピークの出現や消滅などを考慮しなければならず、これらを考慮しつつ、高精度でピッチを検出するのは困難である。

【0023】

また、ピークという極大値を扱う性質上、雑音に対する耐性もあまり期待できない。さらに、時間的な変動を評価するために、フレーム間振幅差分調波性評価部104においては、フレーム間差分に対して調波構造性の有無を評価しているが、単に、振幅の差分を用いているため、調波構造の有する情報が失われてしまうだけではなく、例えば突発雑音が生じた場合には、差分値として突発雑音の音響特徴量がそのまま評価されてしまうという問題がある。

【0024】

そこで、本発明は上述の課題を解決するためになされたものであり、入力信号のレベル変動に依存せず、精度良く音声区間を検出可能な音声区間検出方法および装置を提供することを目的とする。

また、リアルタイム性に優れた音声区間検出方法および装置を提供することも目的とする。

【0025】

【課題を解決するための手段】

本発明のある局面に係る音声区間検出方法は、入力音響信号から音声が含まれる区間を音声区間として検出する音声区間検出方法であって、前記入力音響信号に対し、あらかじめ定められた時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出ステップと、前記音響特徴量のフレーム間における持続性を評価し、評価結果に従って音声区間を決定する区間決定ステップとを含むことを特徴とする。

【0026】

このように、音響特徴量のフレーム間における持続性を評価することにより、音声区間の決定を行なっている。このため、ローカルピークをトラッキングする

従来の方法のようにローカルピークの出現や消滅など、入力信号のレベル変動を考慮する必要がなく、精度よく音声区間を決定することができる。

【0027】

好ましくは、前記音響特徴量抽出ステップでは、前記入力音響信号に対しフレーム単位で周波数変換を行ない、前記周波数変換の結果より調波構造のみを強調し、前記音響特徴量を抽出することを特徴とする。

【0028】

音声（特に母音）には、調波構造性が見られる。このため、調波構造を強調した音響特徴量を用いて音声区間を決定することにより、さらに精度よく音声区間を決定することができる。

【0029】

さらに好ましくは、前記音響特徴量抽出ステップでは、さらに、前記周波数変換の結果より調波構造を抽出し、当該調波構造を含む所定の帯域の周波数変換の結果を、前記音響特徴量とすることを特徴とする。

【0030】

調波構造性が保たれている帯域のみからなる音響特徴量を用いて音声区間を決定することにより、さらに精度よく音声区間を決定することができる。

【0031】

さらに好ましくは、前記区間決定ステップでは、前記音響特徴量のフレーム間における相関値に基づいて、前記持続性を評価することを特徴とする。

【0032】

このように、調波構造の持続性をフレーム間の音響特徴量の相関値により評価している。このため、フレーム間での振幅差分を取り調波構造の持続性を評価する従来方法に比べ、調波構造の有する情報を残した評価が可能である。よって、短いフレームにわたる突発雑音が生じたような場合であっても、そのような突発雑音を音声区間として検出することがなくなり、精度よく音声区間を決定することができる。

【0033】

さらに好ましくは、前記区間決定ステップは、前記音響特徴量のフレーム間に

における持続性を評価する評価値を算出する評価ステップと、前記評価値の時間的な連続性を評価し、評価結果に従って音声区間を決定する音声区間決定ステップとを含むことを特徴とする。

【0034】

音声区間決定ステップでの処理は、実施の形態に述べるように、時間的に連続する有声区間（評価値のみから求められた音声区間）を連結して音声区間を検出する処理に相当する。このように、時間的に連続する有声区間を連結し、音声区間を決定することにより、母音に比べ調波構造性が小さい子音をも音声区間と決定することができる。

【0035】

さらに好ましくは、前記区間決定ステップは、さらに、所定数のフレームにわたる前記評価ステップにおいて算出される前記評価値と第1の所定しきい値との比較に基づいて、前記入力音響信号の音声雑音比を推定するステップと、推定された前記音声雑音比が第2の所定しきい値以上の場合には、前記評価ステップにおいて算出される前記評価値に基づいて前記音声区間を決定するステップとを含み、前記音声区間決定ステップでは、前記音声雑音比が前記第2の所定しきい値未満の場合に、前記評価値の時間的な連続性を評価し、評価結果に従って前記音声区間を決定することを特徴とする。

【0036】

これにより、入力音響信号の音声雑音比が良好な場合には、音響特徴量のフレーム間における持続性を評価する評価値の時間的な連続性を評価し、前記音声区間を決定する処理を省略することができる。このため、リアルタイム性に優れた音声区間の検出が可能になる。

【0037】

なお、本発明は、以上のような音声区間検出方法として実現することができるだけでなく、そのステップを手段とする音声区間検出装置として実現したり、音声区間検出方法の各ステップをコンピュータに実行させるためのプログラムとして実現したりすることもできる。そのようなプログラムは、CD-ROM等の記録媒体やインターネット等の伝送媒体を介して配信することができるのはいうま

でもない。

【0038】

【発明の実施の形態】

(実施の形態1)

以下、図面を参照しながら本発明の実施の形態1に係る音声区間検出装置について説明する。図1は、本実施の形態に係る音声区間検出装置20のハードウェア構成を示すブロック図である。

【0039】

音声区間検出装置20は、入力音響信号（以下、単に「入力信号」という。）の中から人間が発声している区間である音声区間を決定する装置であり、FFT部200と、調波性抽出部201と、有声評価部210と、音声区間決定部205とを備える。

【0040】

FFT部200は、入力信号にFFTを施し、フレームごとにパワースペクトル成分を求める。ここで、1フレームあたりの時間は10msecとするが、この時間に限定されるものではない。

【0041】

調波性抽出部201は、FFT部200で抽出されたパワースペクトル成分から雑音成分等を取り除き、調波構造のみを残したパワースペクトル成分を抽出する。

【0042】

有声評価部210は、調波性抽出部201で抽出された調波構造のフレーム間での相関性を評価することにより、母音の区間であるか否かを評価し、有声区間を抽出する装置であり、特徴量保存部202と、特徴量フレーム間相関値算出部203と、差分処理部204とを備える。なお、調波構造性は、母音の発声区間内のパワースペクトル分布において主に見られる性質であり、子音の発声区間内のパワースペクトル分布においては、母音ほどの調波構造性は見られない。

【0043】

特徴量保存部202は、調波性抽出部201より出力されるパワースペクトル

を所定数のフレーム分保存する。特徴量フレーム間相関値算出部 203 は、調波性抽出部 201 より出力されるパワースペクトルと、特徴量保存部 202 に保存されている一定フレーム前のパワースペクトルとの相関値を算出する。差分処理部 204 は、特徴量フレーム間相関値算出部 203 で求められた相関値のある一定期間における平均値を求め、特徴量フレーム間相関値算出部 203 より出力される相関値から平均値を引き、相関値と平均値との平均差分による補正相関値を求める。

【0044】

音声区間決定部 205 は、差分処理部 204 より出力される平均差分による補正相関値に基づいて、音声区間を決定する。

【0045】

以上のように構成された音声区間検出装置 20 の動作について以下に説明する。図 2 は、音声区間検出装置 20 が実行する処理のフローチャートである。

【0046】

FFT 部 200 は、調波構造を抽出するために使用する音響特徴量として、入力信号に FFT を施すことにより、パワースペクトル成分を求める (S2)。より具体的には、FFT 部 200 は、入力信号を所定のサンプリング周波数 F_s (たとえば、11.025 kHz) でサンプリングし、1 フレーム (たとえば、10 msec) ごとに、所定のポイント (たとえば、1 フレームあたり 128 ポイント) で FFT のスペクトル成分を求める。FFT 部 200 は、各ポイントで求められたスペクトル成分を対数化することによりパワースペクトル成分を求める。以下、パワースペクトル成分を、適宜単にスペクトル成分と表記する。

【0047】

次に、調波性抽出部 201 は、FFT 部 200 で抽出されたパワースペクトル成分から雑音成分等を取り除き、調波構造のみを残したパワースペクトル成分を抽出する (S4)。

【0048】

FFT 部 200 で算出されたパワースペクトル成分には、雑音によるオフセットや声道形状によって形成されるスペクトル包絡形状が含まれており、それぞれ

が時間変動を起こしている。このため、調波性抽出部 201 は、これらの成分を取り除き、声帯振動によって形成される調波構造のみをとりだす。これにより、より効果的に有声区間検出が行なわれる。

【0049】

調波性抽出部 201 による処理 (S4) を図 3 および図 4 を参照しながらより詳細に説明する。図 3 は、調波性抽出部 201 による調波構造抽出処理のフローチャートであり、図 4 は、各フレームにおけるスペクトル成分から調波構造を抽出する過程を模式的に示す図である。

【0050】

図 4 (a) に示されるように、調波性抽出部 201 は、各フレームのスペクトル成分 $S(f)$ より、その極大値をピークホールドした値 $H_{max}(f)$ を算出し (S22)、スペクトル成分 $S(f)$ の極小値をピークホールドした値 $H_{min}(f)$ を算出する (S24)。

【0051】

図 4 (b) に示されるように、調波性抽出部 201 は、スペクトル成分 $S(f)$ から極小値のピークホールド値 $H_{min}(f)$ を引くことにより、スペクトル成分 $S(f)$ に含まれるフロア成分を除去する (S26)。これにより、雑音オフセット成分およびスペクトル包絡に起因する変動成分が除去される。

【0052】

図 4 (c) に示されるように、調波性抽出部 201 は、極大値のピークホールド値 $H_{max}(f)$ と極小値のピークホールド値 $H_{min}(f)$ との差分値を求め、ピーク変動量を算出する (S28)。

【0053】

図 4 (d) に示されるように、調波性抽出部 201 は、ピーク変動量を周波数方向に微分し、その変化量を算出する (S30)。これは、調波構造成分を有する帯域では、ピーク変動量の変化が小さいという仮定に基づいて、調波構造の検出を行うことを目的としている。

【0054】

図 4 (e) に示されるように、調波性抽出部 201 は、上記仮定が反映される

ような重み $W(f)$ を算出する(S32)。すなわち、調波性抽出部201は、ピーク変動量の変化量の絶対値と所定のしきい値とを比較し、当該変化量の絶対値が所定のしきい値 θ 以下であれば重み $W(f)$ を1とし、所定のしきい値 θ 以上であれば当該変化量の絶対値の逆数を重み $W(f)$ とする。これにより、ピーク変動量の変化が大きい部分の重みを小さくし、ピーク変動量の変化が小さい部分の重みを大きくすることができる。

【0055】

図4(f)に示されるように、調波性抽出部201は、フロア成分が除去されたスペクトル成分($S(f) - H_{min}(f)$)に重み $W(f)$ を掛け合わせ、スペクトル成分 $S'(f)$ を求める(S34)。この処理により、ピーク変動量の変化の大きい非調波構造成分を除去することが可能となる。

【0056】

再度、図2に示される音声区間検出装置20の動作説明を続ける。調波構造抽出処理(図2のS4、図3)の後、特徴量フレーム間相関値算出部203は、調波性抽出部201より出力されるスペクトル成分と、特徴量保存部202に保存されている所定フレーム前のスペクトル成分との間の相関値を算出する(S6)。

【0057】

ここでは、着目しているフレームを j 番目のフレームとした場合、隣接するフレームのスペクトル成分を用いて相関値 $E1(j)$ を求める方法について説明する。相関値 $E1(j)$ は、次式(1)～(5)に従い求められる。すなわち、 i フレームおよび $i-1$ フレームの128ポイントにおけるパワースペクトル成分 $P(i)$ および $P(i-1)$ を次式(1)および(2)でそれぞれ表すものとする。また、パワースペクトル成分 $P(i)$ および $P(i-1)$ の相関関数 $\text{xcorr}(P(j-1), P(j))$ の値を次式(3)で表すものとする。すなわち、相関関数 $\text{xcorr}(P(j-1), P(j))$ の値は、各ポイントにおける内積値からなるベクトル量である。 $z1(i)$ を次式(4)に示されるように $\text{xcorr}(P(j-1), P(j))$ のベクトルの要素の最大値を求める。これを j フレームの相関値 $E(j)$ としてもよいし、次式(5)で表されるようにた

例えば3フレーム分加算した値を用いても良い。

【0058】

【数1】

$$P(i) = (p1(i), p2(i), \dots, p128(i)) \quad \dots (1)$$

$$P(i-1) = (p1(i-1), p2(i-1), \dots, p128(i-1)) \quad \dots (2)$$

$$\begin{aligned} \text{xcorr}(P(i-1), P(i)) = \\ (p1(i-1) \times p1(i), p2(i-1) \times p2(i), \dots, p128(i-1) \times p128(i)) \end{aligned} \quad \dots (3)$$

$$z1(i) = \max(\text{xcorr}(P(i-1), P(i))) \quad \dots (4)$$

$$E1(j) = \sum_{i=j-2}^j z1(i) \quad \dots (5)$$

【0059】

相関値 $E(j)$ の一例を図5に示すグラフを用いて説明する。図5は、入力信号を処理することにより得られる信号を表すグラフである。図5(a)は入力信号の波形を示している。この波形は、掃除機の雑音 ($SNR = 0.5 \text{ dB}$) がある環境において、約1200~3000msecの間に「アールアンドビーホテルヒガシニホン」と発音している場合の波形である。この入力信号には、約500msecの箇所に掃除機を動かした際の「カタッ」という突発音が含まれ、2800msec頃に掃除機のモータの回転速度を弱から強に変更し、掃除機の音のレベルが大きくなっている。図5(b)は、図5(a)に示される入力信号にFFTを施した場合のパワーを示しており、図5(c)は、相関値算出処理(S6)で求められた相関値の遷移を示している。

【0060】

ここで、相関値 $E(j)$ の算出は、以下に示すような知見に基づいて算出される。すなわち、フレーム間の音響特徴量の相関値は、時間的に連続するフレーム

において調波構造が連続していることに基づいている。このため、この調波構造を時間的に近いフレーム同士で相関をとることで、有声検出が行なわれる。調波構造が時間的に持続するのは主に母音区間である。このため、母音区間では相関値は大きくなり、子音区間では母音区間よりも小さくなるものと想定される。このように、調波構造に着目しフレーム間でパワースペクトル成分の相関値をとることによって、非周期的な雑音区間においては、相関値が小さくなるものと考えられる。このため、有声区間がより際立って識別可能となる。

【0061】

また、一般的な発話スピードにおいて母音区間の持続時間は50～150 msec（5～15フレーム）と言われており、その持続時間内であれば、フレーム間の相関係数の値は隣接するフレームでなくとも高くなるものと想定できる。この仮定が正しければ、やはり非周期的な雑音の影響を受けにくい評価関数であるということがいえる。相関値 $E1(j)$ を算出する際に、数フレームにわたる相関関数の値の和を用いているのは、突発的に生じる雑音の影響を除去するためと、母音であれば、上記のように50～150 msecの持続時間があるという知見によるものである。従って、図5（c）に示されるように、50フレームの近傍で発声する突発音に対しては反応せずに、相関値は小さいままである。

【0062】

次に、差分処理部204は、特徴量フレーム間相関値算出部203で算出された相関値の一定時間にわたる平均値を求め、各フレームにおける相関値から当該平均値を減算し、平均差分による補正相関値を求める（S8）。なぜならば、相関値から平均値を引くことにより、長時間にわたり生じている周期性の雑音の影響を取り除くことができると考えられるためである。ここでは、5秒程度の相関値の平均値を求めており、図5（c）では、平均値を実線502で示している。すなわち、実線502よりも上の部分に相関値が存在する区間が上記平均差分による補正相関値が正の区間である。

【0063】

次に、音声区間決定部205は、主に有音区間を検出する相関値 $E(j)$ の差分処理部204で算出された平均差分による補正相関値に基づいて、後述する、

相関値による選別、区間の持続長、子音区間や促音区間を加味した区間の連結、の3つの区間補正方法に従い音声区間を決定する（S10）。

【0064】

ここで、音声区間決定部205による音声区間決定処理（図2のS10）についてより詳細に説明する。図6は、一発声単位で音声区間決定する処理の詳細を示すフローチャートである。

【0065】

まず、第一の補正方法である相関値による区間の判定について述べる。音声区間決定部205は、着目しているフレームについて、差分処理部204で求められた補正相関値が所定のしきい値よりも大きいかな否かを調べる（S44）。たとえば、所定のしきい値を0とした場合には、図5（c）に示される相関値が相関値の平均値（実線502）よりも大きいかな否かを調べることに等価である。

【0066】

補正相関値が所定のしきい値よりも大きい場合には（S44でYES）、当該着目フレームは音声フレームであると判断し（S46）、補正相関値が所定のしきい値以下の場合には（S44でNO）、当該着目フレームは非音声フレームであると判断する（S48）。以上の音声判断処理（S44～S48）を音声区間検出対象となっているすべてのフレームについて繰返す（S42～S50）。以上の処理により、図5（d）に示されるようなグラフが得られ、音声フレームが連続する区間が有声区間として検出される。

【0067】

このように、補正相関値の値がしきい値以下である場合には、そのフレームを非音声フレームであると判断する。ただし、騒音のレベルの影響や、音響特徴量のさまざまな条件に応じて、検出区間において期待される補正相関値が異なる。このため、音声フレームと非音声（雑音）フレームとを区別するためのしきい値は、事前の実験を通じて適宜定め用いるものとする。この処理により、中期的な500ms程度の周期雑音を非音声フレームとすることが期待できる。

【0068】

次に、第二の補正方法である隣接有声区間の連結法について述べる。音声区間

決定部 205 は、着目している有声区間と、当該有声区間に隣接する有声区間との間の距離が所定フレーム数未満であるかを調べる (S54)。たとえば、ここでは所定フレーム数を 30 フレームとする。当該距離が 30 フレーム未満の場合には (S54 で YES)、隣接する 2 つの有声区間を連結する (S56)。以上の処理 (S54 ~ S56) をすべての有声区間について行なう (S52 ~ S58)。以上の有声区間連結処理により、図 5 (e) に示されるようなグラフが得られ、近接する有声区間が連結されていることが分かる。

【0069】

有声区間の連結をするのは、以下のような理由による。すなわち、子音区間、特に破裂音 (/k/, /c/, /t/, /p/) や摩擦音などの無声子音の区間においては、調波構造が表れにくいため、相関値が小さく、有声区間として検出されにくい。しかし、子音の近傍には母音が存在するため、母音が連続する区間は有声区間とみなされるという理由による。これにより、子音部分も有声区間とすることが可能になる。

【0070】

最後に、第三の補正方法である区間持続時間について述べる。音声区間決定部 205 は、着目している有声区間について、その持続時間が所定時間よりも長いかな否かを調べる (S62)。たとえば、所定時間は、50 msec であるとする。持続時間が 50 msec よりも長い場合には (S62 で YES)、当該有声区間を音声区間と決定し (S64)、持続時間が 50 msec 以下の場合には (S62 で NO)、当該有声区間を非音声区間と決定する (S66)。以上の処理 (S62 ~ S66) をすべての有声区間について行なうことにより音声区間が決定される (S60 ~ S68)。以上説明した処理により、図 5 (f) に示すようなグラフが得られ、110 ~ 280 フレームあたりに音声区間が検出される。また、図 5 (e) のグラフに存在していた 325 フレームあたりに存在していた周期性ノイズに対する有声区間は、非音声区間と決定されていることが分かる。このように、有声区間の持続時間により有声区間を選別する処理では、相関値が高い短時間の周期的雑音を取り除くことができる。

【0071】

以上説明したように本実施の形態によれば、調波構造を有するスペクトル成分のフレーム間での持続性を評価することにより、有声区間を決定している。このため、ローカルピークをトラッキングする従来の方法に比べ、精度よく音声区間を決定することができる。

【0072】

特に、調波構造の持続性をフレーム間のスペクトル成分の相関値により評価している。このため、フレーム間での振幅差分を取り調波構造の持続性を評価する従来方法に比べ、調波構造の有する情報を残した評価が可能である。よって、短いフレームにわたる突発雑音が生じたような場合であっても、突発雑音を有声区間として検出することがない。

【0073】

また、時間的に隣接する有声区間を連結することにより音声区間と決定している。このため、母音に比べ調波構造性が小さい子音をも音声区間と決定することが可能である。また、有声区間の持続時間を評価することにより、周期性を有する雑音を除去することが可能になる。

【0074】

(実施の形態2)

以下、図面を参照しながら本発明の実施の形態2に係る音声区間検出装置について説明する。本実施の形態に係る音声区間検出装置では、入力信号のSNRがよい場合には、フレーム間でのスペクトル成分の相関性のみから音声区間を決定する点が実施の形態1に係る音声区間検出装置とは異なる。

【0075】

図7は、本実施の形態に係る音声区間検出装置30のハードウェア構成を示すブロック図である。実施の形態1に係る音声区間検出装置20と同一の構成要素については、同一の参照番号を付す。その名称および機能も同一であるため、適宜説明を省略する。

【0076】

音声区間検出装置30は、入力信号の中から人間が発声している区間である音声区間を決定する装置であり、FFT部200と、調波性抽出部201と、20

10と、SNR推定部206と、音声区間決定部205とを備える。

【0077】

有声評価部210は、有声区間を抽出する装置であり、特徴量保存部202と、特徴量フレーム間相関値算出部203と、差分処理部204とを備える。

【0078】

SNR推定部206は、差分処理部204より出力される平均差分による補正相関値に基づいて、入力信号のSNRを推定する。SNR推定部206は、SNRが悪いと推定される場合には、差分処理部204より出力される補正相関値を音声区間決定部205に出力し、SNRがよいと推定される場合には、音声区間決定部205への補正相関値の出力は行わずに、差分処理部204より出力される補正相関値より音声区間を決定する。これは、入力信号のSNRが良好な場合には、音声区間と非音声区間との相関値の差がはっきりとしているという性質があるためである。

【0079】

次に、SNR推定部206による入力信号のSNRの推定方法について説明する。SNR推定部206は、差分処理部204で求められる相関値の平均値が所定のしきい値未満の場合には、SNRが良好であると推定し、当該平均値が所定のしきい値以上の場合には、SNRが悪いと推定する。これは、以下のような理由に基づく。すなわち、相関値の平均値を、一発声の持続時間よりも十分に長い時間（たとえば、5秒間）にわたって求めると、SNRが良好な環境下においては、雑音区間における相関値が小さくなるため、相関値の平均値が小さくなる。これに対し、周期性の雑音を有するようなSNRが悪い環境下においては、雑音区間における相関値が大きくなるため、相関値の平均値が大きくなる。このように、相関値の平均値とSNRとが連動しているという性質を用いることにより、既に計算済みの一つのパラメータを評価するだけで簡単にSNRを推定することが可能である。

【0080】

以上のように構成された音声区間検出装置30の動作について以下に説明する。図8は、音声区間検出装置30が実行する処理のフローチャートである。

FFT部200によるFFT処理(S2)から差分処理部204による補正相関値算出処理(S8)までは、図2に示した実施の形態1における音声区間検出装置20の動作と同様である。そのため、その詳細な説明はここでは繰返さない。

【0081】

次に、SNR推定部206は、上記方法に従い、入力信号のSNRを推定する(S12)。SNRが良好であると推定される場合には(S14でYES)、所定のしきい値を超える補正相関値を音声区間として決定する(S16)。SNRが悪いと推定される場合には(S14でNO)、図2および図6を参照して説明した実施の形態1に係る音声区間決定部205による音声区間決定処理(図2のS10)と同様の処理を実行し、音声区間を決定する(S10)。

【0082】

以上説明したように、本実施の形態によると、実施の形態1に記載の効果に加え、入力信号のSNRが良好な場合には、有声区間の連続性および持続時間による音声区間決定処理を行なう必要がなくなる。このため、リアルタイム性に優れた音声区間の検出が可能になる。

【0083】

このように、実施の形態1および2によると、環境雑音下で発声を行なった場合であっても、入力信号のレベル変動に依存せず、精度よく音声区間を決定することができる。また、突発雑音や周期性雑音の影響を取り除き、精度良く音声区間を検出することができる。さらに、リアルタイムで音声区間を検出することができる。さらにまた、調波構造的な小さい子音部分をも音声区間として精度良く検出することができる。また、入力信号を周波数変換したスペクトル成分にローカットフィルタをかけることにより、スペクトル包絡成分を除去することができる。

【0084】

以上、本発明に係る音声区間検出装置について実施の形態1および実施の形態2に基づいて説明したが、本発明はこれらの実施の形態に限定されるものではない。

【0085】

(FFT部200の変形例)

たとえば、上述の実施の形態では、音響特徴量としてFFTパワースペクトル成分を用いる方法について述べたが、FFTスペクトル成分そのものや、フレーム長単位の自己相関関数や、時間軸上での線形予測残差のFFTパワースペクトル成分を用いてもよい。また、FFTスペクトルからFFTパワースペクトルを求める前に、各スペクトル成分を二乗するなどの方法により、極大値および極小値の差を拡大させ、調波構造を強調させてもよい。さらに、FFTスペクトルの対数を取り、FFTパワースペクトルを求める代わりに、FFTスペクトルの平方根を求め、FFTパワースペクトルとしてもよい。さらにまた、FFTスペクトル成分を求める前に、時間軸データに対して、フレームごとにハミング窓などの係数をかけてもよいし、プリアンファシス処理 ($1-z^{-1}$) を行うことで、高域強調を行ってもよい。また、音響特徴量として線スペクトル周波数 (LSF) を用いてもよい。また、周波数変換演算として、FFTに限られるものではなく、DFT (Discrete Fourier Transform)、DCT (Discrete Cosine Transform)、DST (Discrete Sine Transform) を用いても良い。

【0086】

(調波性抽出部201の変形例)

また、調波性抽出部201によるスペクトル成分 $S(f)$ に含まれるフロア成分の除去処理 (図3のS26) の代わりに、スペクトル成分 $S(f)$ にローカットフィルタを通過させるようにしてもよい。各フレームのスペクトル成分 $S(f)$ を周波数軸方向に並べた波形とみなすと、スペクトル包絡成分は、調波構造に比べゆっくりした変動である。このため、スペクトル成分にローカットフィルタをかけることにより、スペクトル包絡成分を除去することができる。この手法は時間軸上でローカットフィルタを用いて低周波数成分を取り除くことに相当するが、帯域パワーやスペクトル包絡などの情報と調音構造とを同時に評価することができる点において、周波数軸上で処理する方法の方が好ましいといえる。ただし、このようなローカットフィルタを用いて算出されたスペクトル成分は、調音構造に起因する変動の他に、非周期雑音や電子音などの単一周波数を有する音声

以外の音を含んでいる可能性がある。しかし、これらの音は、有声評価部 210 や音声区間決定部 205 の処理により除去される。

【0087】

その他のフロア成分除去の方法としては、各スペクトル成分のうち、所定の基準値以下のスペクトル成分は利用しないようにする方法がある。基準値の算出方法としては、全フレームのスペクトル成分の平均値を基準値に用いる方法、一発声の持続時間よりも十分に長い時間（たとえば、5 秒間）におけるスペクトル成分の平均値を基準値に用いる方法、スペクトル成分をいくつかの帯域に予め分割しておき、帯域ごとにスペクトル成分の平均値を求める基準値とする方法などがある。特に、静かな環境からうるさい環境へ変化するなどの環境の変動がある場合には、基準値として、全フレームのスペクトル成分の平均値を利用するよりも、現在検出しようとしているフレームを含む数秒程度の区間のスペクトル成分の平均値を用いるのがよい。

【0088】

（特徴量フレーム間相関値算出部 203 の変形例）

また、特徴量フレーム間相関値算出部 203 は、相関関数として、式（3）の代わりに、次式（6）を用いて相関値 $E1(j)$ を求めるようにしてもよい。ここで、式（6）は、 $P(i-1)$ および $P(i)$ を 128 次元ベクトル空間中のベクトルとした場合の 2 つのベクトル $P(i-1)$ および $P(i)$ がなす角の余弦を示している。また、特徴量フレーム間相関値算出部 203 は、相関値 $E1(j)$ の代わりにフレーム j と 4 フレーム離れたフレーム間相関値を特徴とさせて、次式（7）および（8）に従い相関値 $E2(j)$ を求めるようにしてもよいし、8 フレーム離れたフレーム間相関値を特徴として、次式（9）および（10）に従い相関値 $E3(j)$ を求めるようにしてもよい。このように、離れたフレーム間で相関値を求めることにより、突発的な環境雑音に強い相関値を得ることができるという特徴がある。

【0089】

さらに、次式（11）～（13）に従い、相関値 $E1(j)$ 、相関値 $E2(j)$ 、相関値 $E3(j)$ の大小関係に応じた相関値 $E4(j)$ を求めるようにして

もよいし、次式(14)に従い相関値 $E1(j)$ 、相関値 $E2(j)$ 、相関値 $E3(j)$ を加算した相関値 $E5(j)$ を求めるようにしてもよいし、次式(15)に従い、相関値 $E1(j)$ 、相関値 $E2(j)$ 、相関値 $E3(j)$ のうちの最大値を相関値 $E6(j)$ を求めるようにしてもよい。

【0090】

【数 2】

$$\begin{aligned} \text{xcorr}(P(i-1), P(i)) &= \frac{P(i-1) \cdot P(i)}{|P(i-1)| |P(i)|} \\ &= \frac{p1(j-1) \times p1(j) + p2(j-1) \times p2(j) + \dots + p128(j-1) \times p128(j)}{\sqrt{p1(j-1)^2 + p2(j-1)^2 + \dots + p128(j-1)^2} \sqrt{p1(j)^2 + p2(j)^2 + \dots + p128(j)^2}} \end{aligned}$$

… (6)

$$z2(i) = \max(\text{xcorr}(P(i-4), P(i))) \quad \dots (7)$$

$$E2(j) = \sum_{i=j-2}^j z2(i) \quad \dots (8)$$

$$z3(i) = \max(\text{xcorr}(P(i-8), P(i))) \quad \dots (9)$$

$$E3(j) = \sum_{i=j-2}^j z3(i) \quad \dots (10)$$

$$E4(j) = z1(j) \quad \dots (11)$$

$$\text{if } (z3(j) > 0.5) \ E4(j) = E4(j) + z1(j)/z3(j) \quad \dots (12)$$

$$\text{if } (z2(j) > 0.5) \ E4(j) = E4(j) + z1(j)/z2(j) \quad \dots (13)$$

$$\begin{aligned} E5(j) &= E1(j) + E2(j) + E3(j) \\ &= \sum_{i=j-2}^j z1(i) + \sum_{i=j-2}^j z2(i) + \sum_{i=j-2}^j z3(i) \end{aligned} \quad \dots (14)$$

$$\begin{aligned} E6(j) &= \max(E1(j), E2(j), E3(j)) \\ &= \max\left(\sum_{i=j-2}^j z1(i), \sum_{i=j-2}^j z2(i), \sum_{i=j-2}^j z3(i)\right) \end{aligned} \quad \dots (15)$$

【0091】

なお、相関値は、上述の $E1(j) \sim E6(j)$ の6つに限定されるわけではなく、これらの相関値を組み合わせて、新たな相関値を算出するようにしてもよい。たとえば、過去に推定された入力音響信号の SNR から、SNR が小さい場

合には、相関値 E (1) を使用し、S N R が大きい場合には、相関値 E (2) または E (3) を使用するようにしてもよい。

【0092】

(音声区間決定部 205 の変形例)

図 6 を用いて説明した音声区間決定部 205 の処理は、相関値による有声区間決定処理 (S 42 ~ S 50)、有声区間の連結処理 (S 52 ~ S 58)、および有声区間の持続時間による音声区間決定処理 (S 60 ~ S 68) の 3 つの処理に大きく分類されるが、これら 3 つの処理を図 6 に示される順序で実行する必要はなく、他の順序で実行するようにしてもよい。また、3 つの処理のうち、1 つまたは 2 つの処理のみを実行するようにしてもよい。また、図 6 は、一発声単位で処理を行う例であるが、たとえば注目フレームごとに相関値による有声区間決定処理のみを行うことで、フレーム単位で音声区間を決定補正してもよい。さらに、リアルタイム性が要求されることを想定して、フレーム単位の相関値による音声区間を速報値として出力しておき、別途、定期的に、一発声等長い単位で補正決定された音声区間を確定値として出力することで、リアルタイム性にも、検出区間性能にも対応可能な、音声検出器として作用させてもよい。

【0093】

(S N R 推定部 206 の変形例)

また、S N R 推定部 206 は、入力信号から直接 S N R を推定するようにしてもよい。たとえば、差分処理部 204 で算出された補正相関値が正の部分 S (シグナル) 部分とし、S 部分のパワーを求め、補正相関値が負の部分 N (ノイズ) 部分とし、N 部分のパワーを求め、S N R を求めるようにする。

【0094】

(その他の変形例)

さらに、上述の音声区間検出処理を前処理とし、音声区間のみについて音声認識を行なう音声認識装置に音声区間検出装置を使用してもよい。

【0095】

また、上述の音声区間検出処理を前処理として、音声区間のみについて録音を行なう I C (Integrated Circuit) レコーダなどの音声録音装置に音声区間検出

装置を使用しても良い。このように、音声区間のみを録音することにより、ICレコーダの記憶領域を効率的に利用することが可能となる。

【0096】

また、音声区間以外の区間の入力信号をカットして雑音を抑制する雑音抑制装置に音声認識装置を利用してもよい。

【0097】

さらにまた、VTR (Video Tape Recorder) 等で撮影された映像から、音声区間の映像を抽出するのに、上述の音声区間検出処理を用いてもよく、映像を編集するオーサリングツールなどにも適用可能である。

【0098】

また、図4 (f) に示されるパワースペクトル成分 $S'(f)$ のうち、調波構造性が最もよく保たれている帯域を1つ以上抽出し、その帯域のみを用いて処理を行うようにしてもよい。

【0099】

【発明の効果】

以上のように、本発明に係る音声区間検出方法および装置によると、音声区間と雑音区間との精度良い選別が可能となり、特に、音声認識方法の前処理として本発明を適用することにより、音声認識率を向上させることができ、その実用的価値は極めて高い。また、IC (Integrated Circuit) レコーダなどに使用することにより音声区間のみを録音したりすることにより、記録容量の効率利用も可能である。

【図面の簡単な説明】

【図1】 本発明の実施の形態1に係る音声区間検出装置のハードウェア構成を示すブロック図である。

【図2】 実施の形態1に係る音声区間検出装置が実行する処理のフローチャートである。

【図3】 調波性抽出部による調波構造抽出処理のフローチャートである。

【図4】 各フレームにおけるスペクトル成分から調波構造を抽出する過程を模式的に示す図である。

【図 5】 本発明による入力信号の変換の遷移を示す図である。

【図 6】 音声区間決定処理のフローチャートである。

【図 7】 本発明の実施の形態 2 に係る音声区間検出装置のハードウェア構成を示すブロック図である。

【図 8】 実施の形態 2 に係る音声区間検出装置が実行する処理のフローチャートである。

【図 9】 従来の音声区間決定装置の概略のハードウェア構成を示す図である。

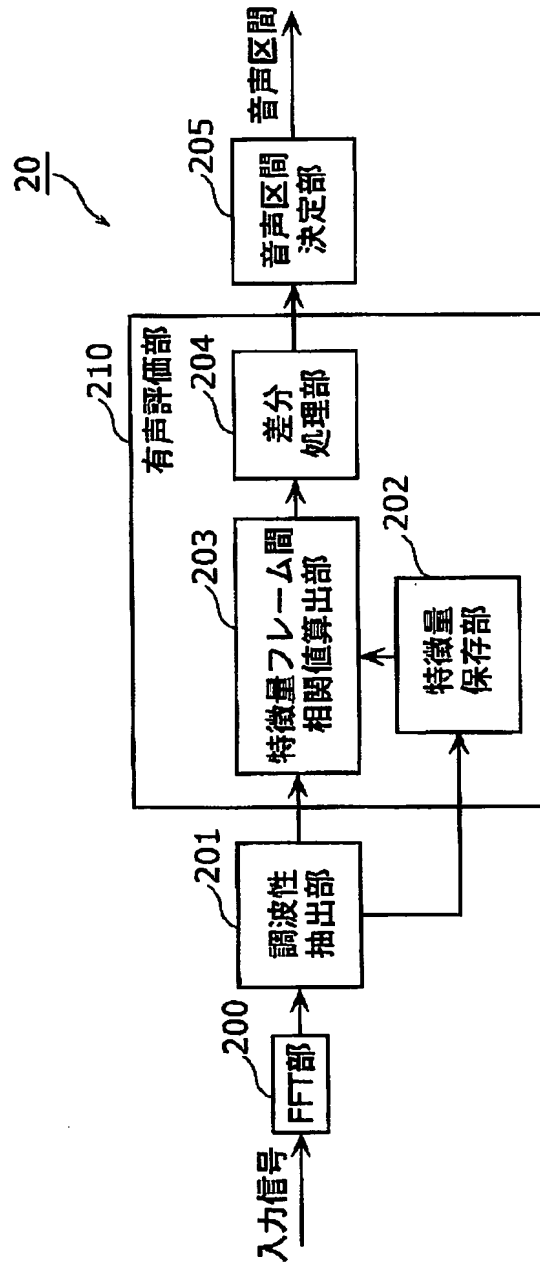
【符号の説明】

- 10, 20, 30 音声区間検出装置
- 100, 200 FFT部
- 101 調波構造評価部
- 102 調波性ピーク検出部
- 103 ピッチ候補検出部
- 104 フレーム間振幅差分調波性評価部
- 105, 206 音声区間決定部
- 201 調波性抽出部
- 202 特徴量保存部
- 203 特徴量フレーム間相関値算出部
- 204 差分処理部
- 206 SNR推定部
- 210 有声評価部

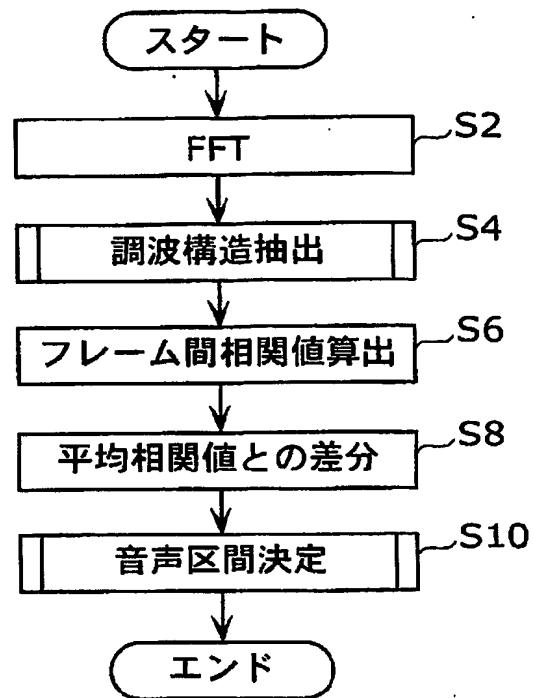
【書類名】

図面

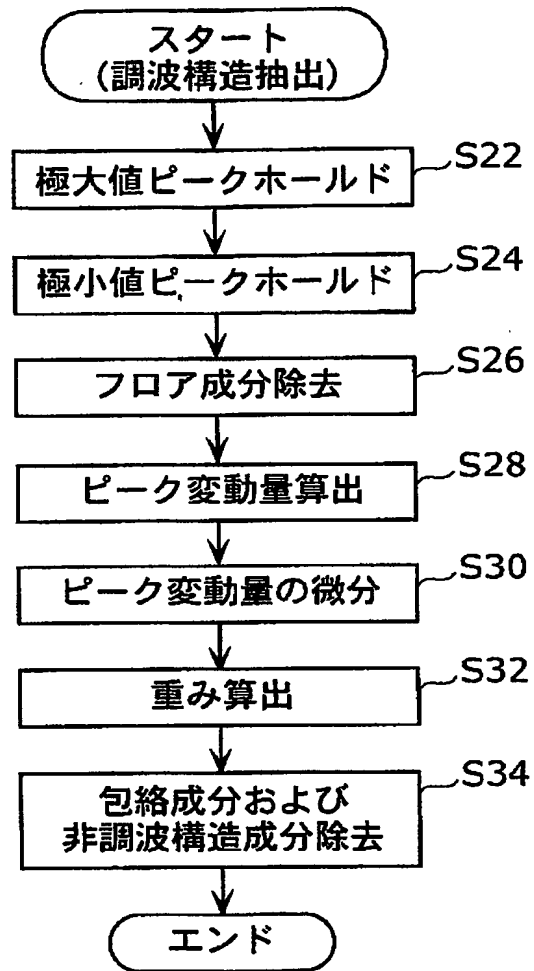
【図 1】



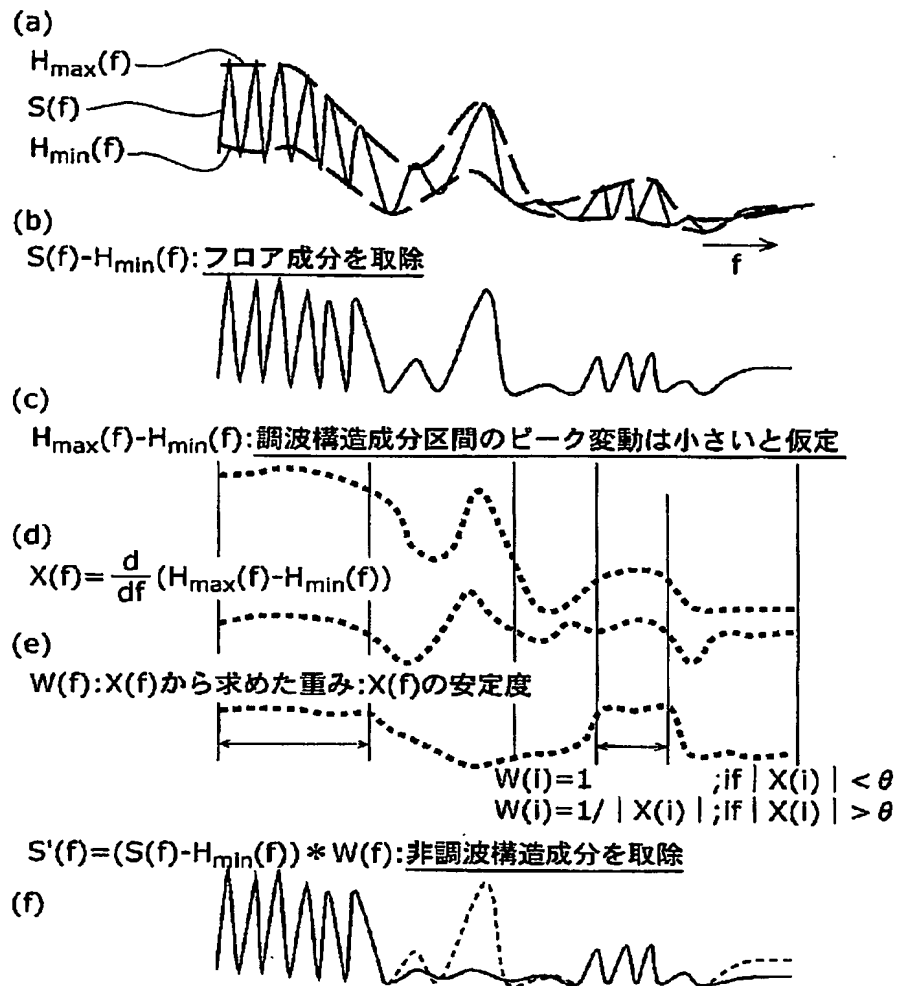
【図 2】



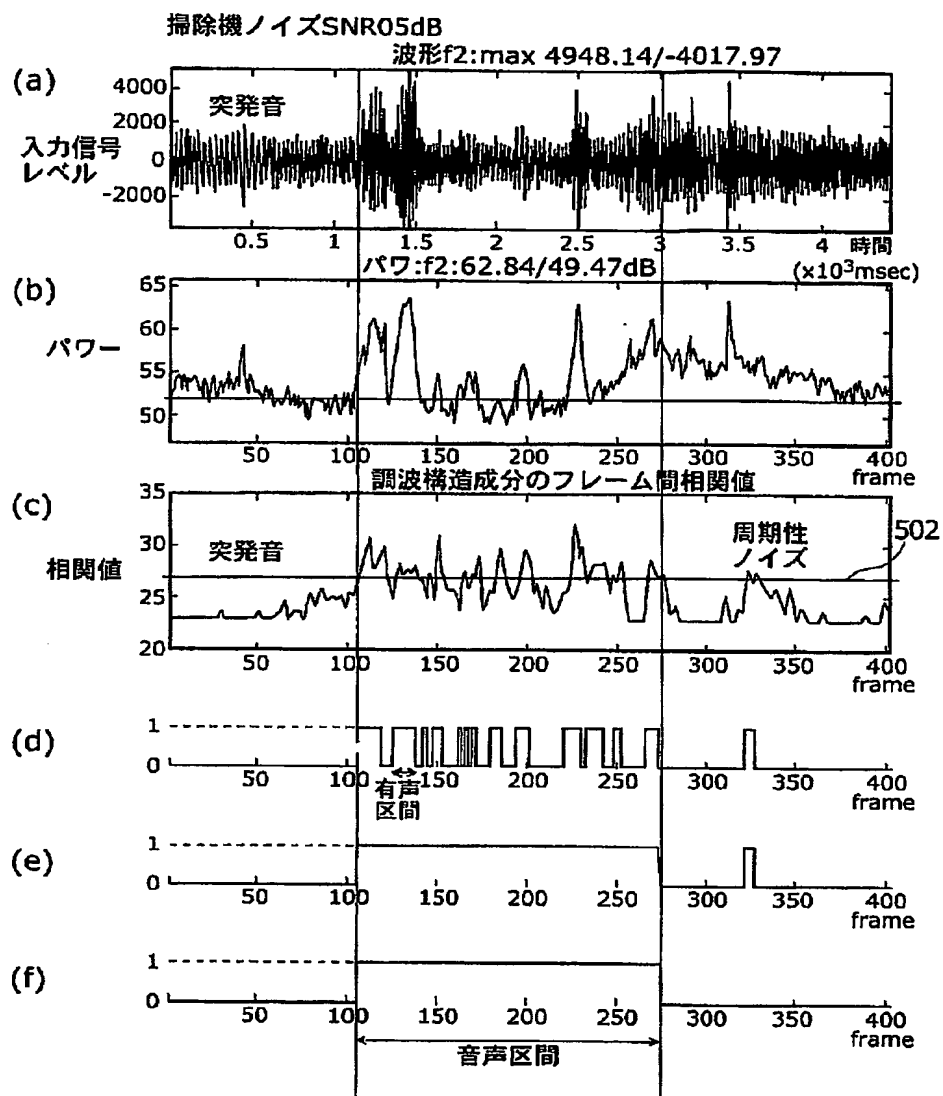
【図 3】



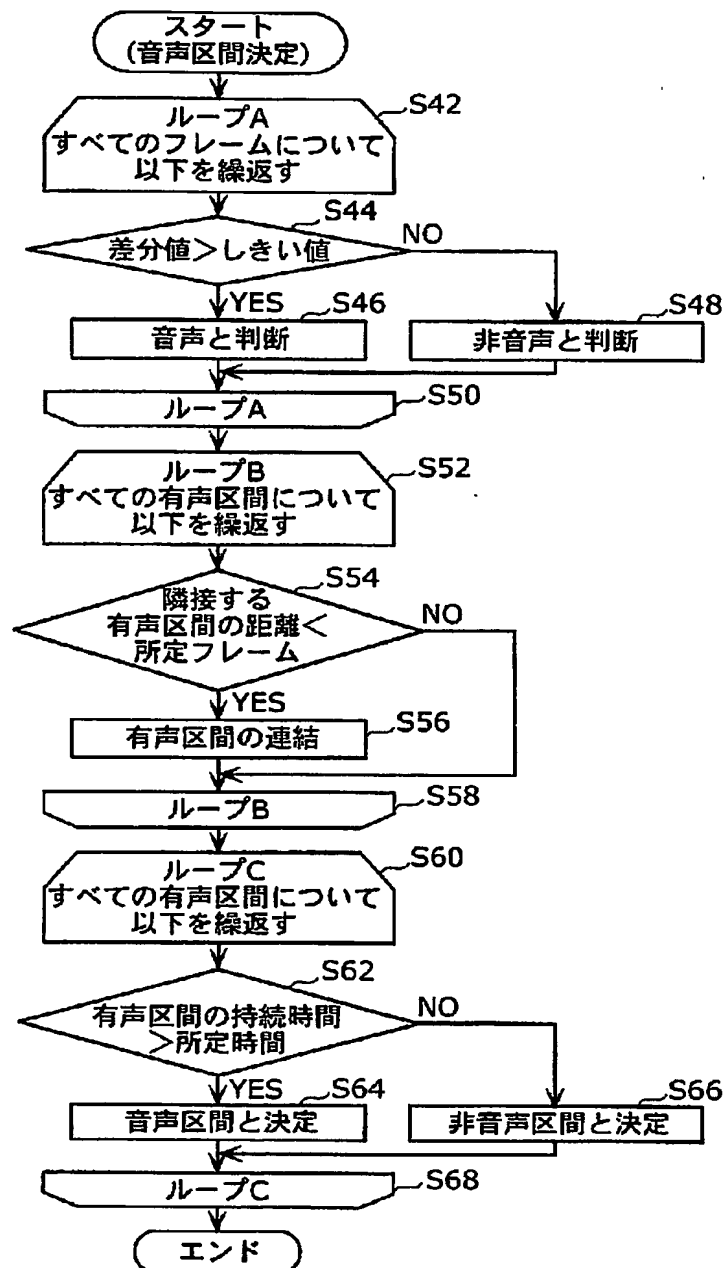
【図 4】



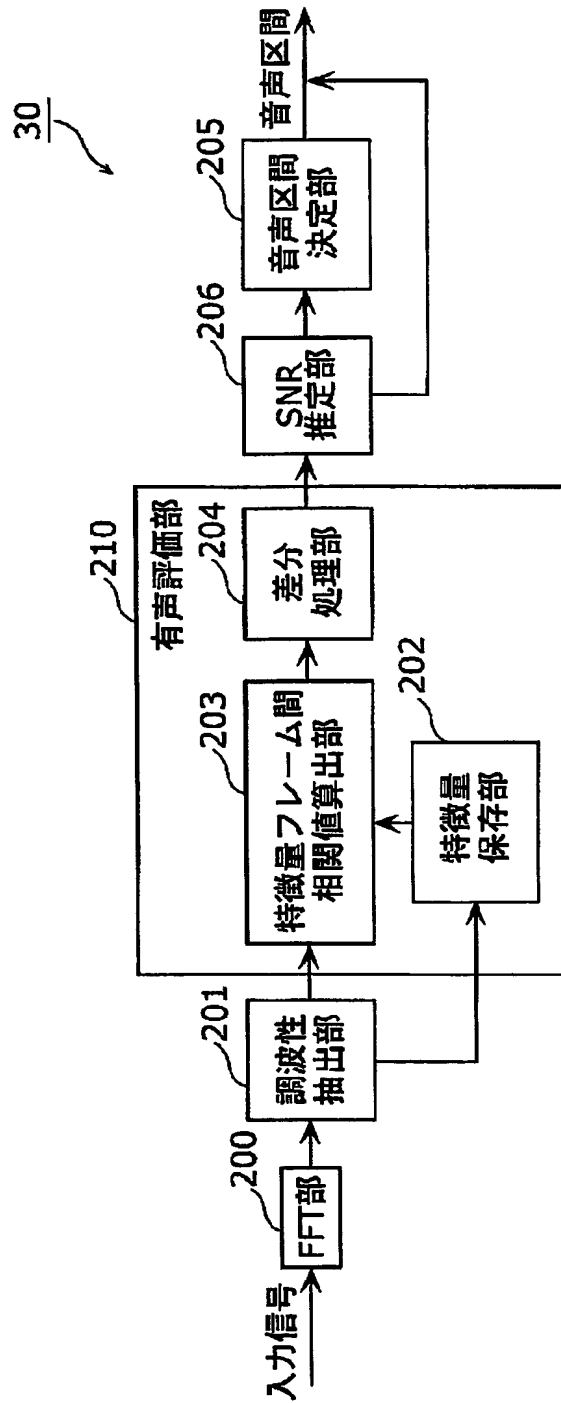
【図 5】



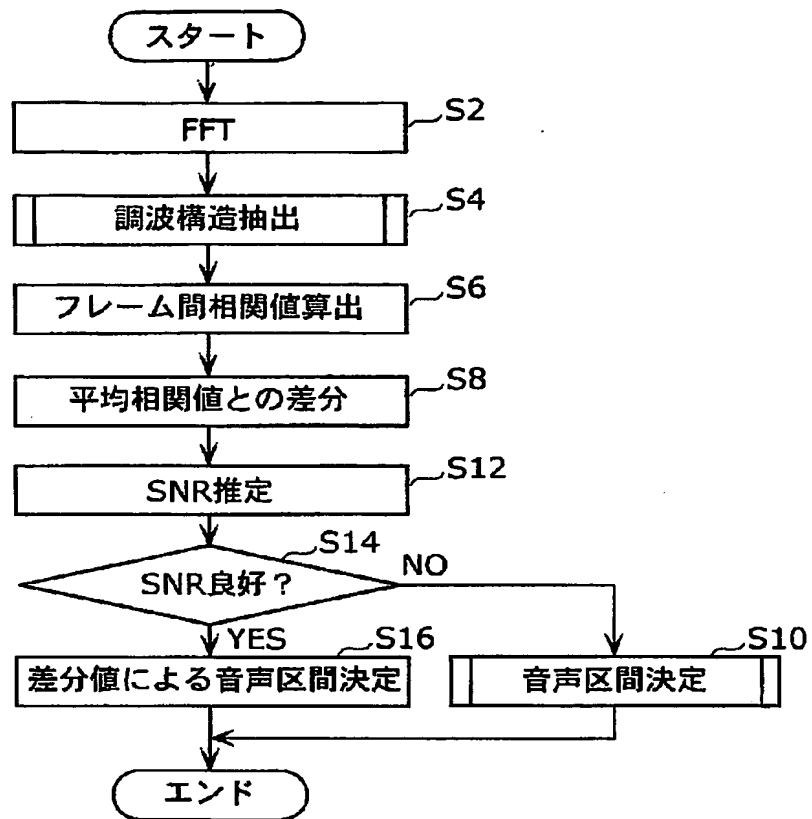
【図 6】



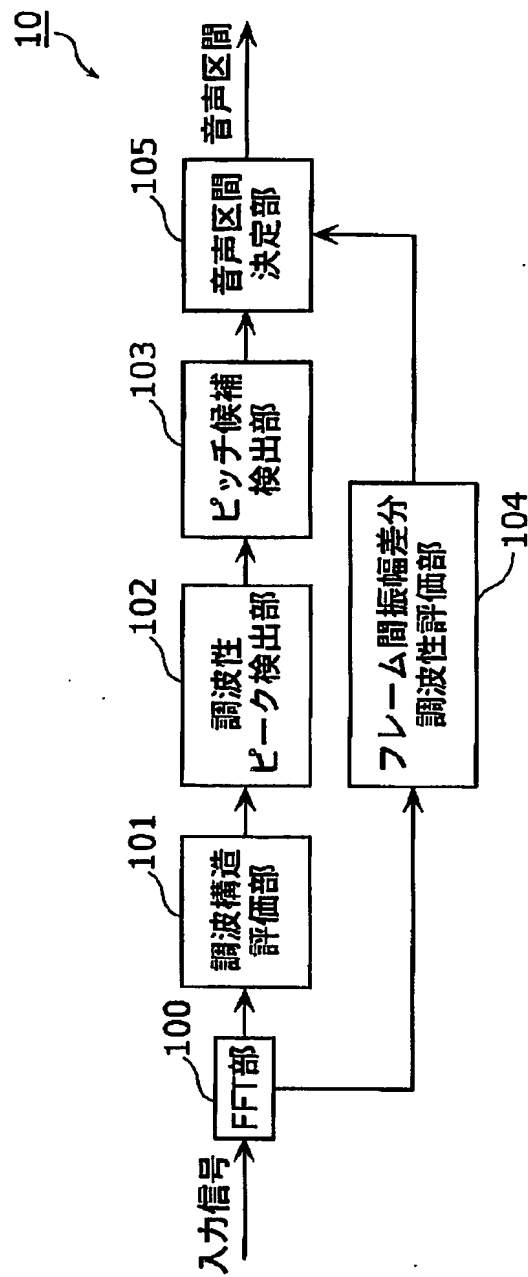
【図 7】



【図 8】



【図 9】



【書類名】 要約書

【要約】

【課題】 入力信号のレベル変動に依存せず、リアルタイム性に優れ、耐雑音性のある音声区間検出装置を提供する。

【解決手段】 入力信号にFFTを施し、フレームごとにパワースペクトル成分を求めるFFT部200と、パワースペクトル成分より調波構造のみを残す調波性抽出部201と、調波性抽出部201で抽出された調波構造のフレーム間での相関性を評価することにより、母音の区間であるか否かを評価し、有声区間を抽出する有声評価部210と、有声評価部210の出力の連続性および持続性に基づいて音声区間を決定する音声区間決定部205とを備える。

【選択図】 図1

認定・付加情報

特許出願の番号	特願 2003-165946
受付番号	50300973606
書類名	特許願
担当官	第八担当上席 0097
作成日	平成15年 6月12日

<認定情報・付加情報>

【提出日】 平成15年 6月11日

次頁無

特願 2003-165946

ページ: 1/E

出願人履歴情報

識別番号

[000005821]

1. 変更年月日

1990年 8月28日

[変更理由]

新規登録

住所

大阪府門真市大字門真1006番地

氏名

松下電器産業株式会社